PATENT APPLICATION
Atty. Docket No.: MDS-020 (6219/6)

SYSTEM FOR NORMALIZING SPECTRA

Cross-Reference to Related Applications

This application is related to the U.S. patent application entitled, "Spectral Data Classification Of Samples" and identified by Attorney Docket Number MDS-021, filed on even date herewith, and the U.S. patent application entitled, "A Spectroscopic System Employing A Plurality of Data Types" and identified by Attorney Docket Number MDS-022, filed on even date herewith. All of the above applications are assigned to the common assignee of this application, and are hereby incorporated by reference.

Government Rights

This invention was made with government support under Contract No. CA66481 awarded by National Cancer Institute, NIH. The government may have certain rights in the invention.

Field of the Invention

This invention relates generally to spectral analysis. More particularly, in one embodiment, the invention relates to analysis of optical spectra using only a portion of the spectral information for data normalization.

Background of the Invention

In general, spectra are recorded as values of amplitude, typically measured as a response to an excitation, as a function of wavelength (or the inverse of wavelength, namely frequency). In the field of spectral analysis, it is often necessary to calibrate or preprocess one or more spectra in order to be able to compare spectra or to extract information from spectra. One

calibration or preprocessing approach is to normalize a spectrum or a set of spectra. Normalization may be required, for example, when comparing spectra having different amplitudes. In the case of optical spectra in particular, differences in amplitude may result from differences in a level of illumination, differences in a response of a detector, or differences in optical behavior of one sample as compared to another. Normalization is a process whereby the differences in instrument performance from spectrum to spectrum are reduced or eliminated.

Two common methods for normalizing spectral information are to normalize a spectrum to a maximum value of amplitude in the spectrum ("peak normalization"), and to normalize a spectrum to an area determined by integrating the spectrum over a range of wavelengths or frequencies ("area normalization"). Peak normalization is performed by dividing the amplitude at each point in a spectrum by the maximum amplitude of that individual spectrum. One obtains a normalized spectrum having intensities ranging from 1.0 at the location of the maximum to possibly as little as 0.0 where the spectral amplitude vanishes. Peak normalization in principle removes the variations in instrument behavior from spectrum to spectrum. However, peak normalization discards information about differences in samples that cause differences in amplitude of response to an invariant excitation. Such information can be very useful, but it is eliminated by normalizing all spectra in a set to a common maximum normalized amplitude of 1.0.

Peak normalization is based on a single amplitude value that appears in a spectrum. To the extent that this single value is incorrect, through a change in illumination intensity, instrumental misalignment, excessive noise in the data, or the like, the peak normalization method will give erroneous information.

Area normalization is another method of normalizing spectra in which the area under the spectrum is computed, for example by integrating the amplitude of the spectrum as a function of wavelength or frequency, and the entire spectrum is recomputed by dividing each value of amplitude by the value determined for the area. The resulting area normalized spectrum has an area of one area unit. However, the energy carried by electromagnetic radiation is proportional to the frequency, ν, of the radiation (e.g., Energy = $h\nu$), or equivalently, is inversely proportional to wavelength, λ, (i.e., Energy = hc/λ), where h is Planck's constant, and c is the speed of light. Therefore, an integration of amplitude over wavelength applies an equal "weight" to a unit of amplitude at long wavelength (i.e., low energy) as a unit of amplitude at short wavelength (i.e., high energy), even though one region may have a far different influence or effect than another, based on the energy content of the radiation.

Summary of the Invention

The invention overcomes the disadvantages of the normalization methods that exist in the prior art, and provides an improved method and system for normalizing spectra. Rather than depending on a single observation, or on the entire range of observations, in a spectrum, in one embodiment, the invention uses as a basis for normalizing the spectrum, the range or ranges of observations within the spectrum that correspond to meaningful content in the spectrum. In one embodiment, the process of the invention is referred to as non-uniform segment normalization because it relies on the use of one or more segments of a spectrum that are not constrained to be of uniform width within the spectrum, nor do the observations have to be evenly spaced in wavelength across the entire spectrum.

In one aspect, the invention features a method of performing spectral analysis. The method includes obtaining an optical spectrum, and normalizing the optical spectrum by application of non-uniform segment normalization. In one embodiment, the method further includes selecting one or more segments from the optical spectrum, each of the segments being bounded by an upper wavelength and a lower wavelength and containing one or more wavelengths, each of the wavelengths having an associated amplitude; determining an area under a curve associated with each particular segment, wherein each the curve is bounded along a first axis by the upper wavelength and the lower wavelength of the particular segment, and along a second axis by the amplitudes associated with each of the wavelengths included in the particular segment; summing the areas for each of the segments to determine a normalization factor; and dividing at least one associated amplitude for one of the wavelengths included in the segments by the normalization factor.

In one embodiment, a first segment differs in size of wavelength range from that of a second segment, wherein the size of wavelength range is defined as the absolute magnitude of a difference between the upper wavelength and the lower wavelength. In another embodiment, a first segment is equal in wavelength range to a second segment.

In one embodiment, the one or more segments comprises at least first and second non-contiguous segments. In another embodiment, the one or more segments comprises at least first, second and third segments, and there exists a first span between an upper wavelength of the first segment and a lower wavelength of the second segment, and a second span between an upper wavelength of the second segment and a lower wavelength of the third segment. In one

embodiment, the first and the second spans differ in magnitude. In an alternative embodiment, the first and the second spans are substantially equal in magnitude.

In one embodiment, the method further comprises obtaining the spectrum from a specimen of human cervical tissue. In another embodiment, the method further comprises extracting a test parameter from the optical spectrum. In still another embodiment, the method further comprises determining a disease status of the test specimen by analyzing the optical spectrum subsequent to the normalizing.

In another aspect, the invention features a system for performing spectral analysis. The system includes a spectrographic device, adapted to obtain an optical spectrum from a test specimen, and a processor adapted to normalize the optical spectrum by application of non-uniform segment normalization.

In one embodiment, the system further includes machine readable instructions executing on the processor and adapted to select one or more segments from the optical spectrum, each of the segments being bounded by an upper wavelength and a lower wavelength, and containing one or more wavelengths, each of the wavelengths having an associated amplitude; determine an area under a curve associated with each particular one of the segments, wherein each the curve is bounded along a first axis by the upper wavelength and the lower wavelength of the particular segment, and along a second axis by the amplitudes associated with each of the wavelengths included in the particular segment; sum the areas for each of the segments to determine a normalization factor; and divide at least one the associated amplitude for one of the wavelengths included in the segments by the normalization factor.

In one embodiment, the machine readable instructions are further adapted to one of select and enable a user to select a first segment different in size of wavelength range from that of a second segment, wherein the size of wavelength range is defined as the absolute magnitude of a difference between the upper wavelength and the lower wavelength. In another embodiment, the machine readable instructions are further adapted to one of select and enable a user to select a first segment to be substantially equal in size of wavelength range to that of a second segment. In still another embodiment, the machine readable instructions are further adapted to one of select and enable a user to select the one or more segments to include at least first and second non-contiguous segments. In yet another embodiment, the machine readable instructions are further adapted to one of select and enable a user to select the one or more segments to include at least first, second and third segments, and there exists a first wavelength span between an upper wavelength of the first segment and a lower wavelength of the second segment, and a second wavelength span between an upper wavelength of the second segment and a lower wavelength of the third segment. In one embodiment, the first and the second spans differ in magnitude. In an alternative embodiment, the first and the second spans are substantially equal in magnitude.

In another embodiment, the spectrographic device is further adapted to obtain the optical spectrum from a specimen of human cervical tissue. In another embodiment, the machine readable instructions are further adapted to extract a test parameter from the optical spectrum. In yet another embodiment, the machine readable instructions are further adapted to determine a disease status of the test specimen by analyzing the optical spectrum, subsequent to the normalizing.

The foregoing and other objects, aspects, features, and advantages of the invention will become more apparent from the following description and from the claims.

Brief Description of the Drawings

The objects and features of the invention can be better understood with reference to the drawings described below, and the claims. The drawings are not necessarily to scale, emphasis instead generally being placed upon illustrating the principles of the invention. In the drawings, like numerals are used to indicate like parts throughout the various views.

FIG. 1 shows an exemplary spectroscopic system employing a non-uniform segment normalization method according to an illustrative embodiment of the invention;

FIG. 2 shows an exemplary operational block diagram of the spectroscopic system of Fig. 1;

FIG. 3 is a schematic flow diagram of an illustrative spectral analysis process incorporating features of the invention;

FIG. 4 is a more detailed schematic flow diagram depicting an exemplary calibration step of the type depicted in FIG. 3 and incorporating a non-uniform segment normalization method according to an illustrative embodiment of the invention;

FIG. 5 is a detailed flow diagram that shows the steps of an exemplary non-uniform segment normalization method, according to an illustrative embodiment of the invention; and

FIG. 6 is a graph depicting a selection of particular wavelength regions for applying a non-uniform segment normalization process according to an illustrative embodiment of the invention.

Detailed Description

The invention will be described in terms of embodiments that relate to the normalization of optical spectra, particularly in the area of medical diagnostics, and especially as it relates to the analysis of spectra obtained from human cervical tissue in the detection of cervical cancer. However, the invention has applicability generally in the area of normalization of optical spectra.

FIG. 1 depicts an exemplary spectroscopic system 100 employing a non-uniform segment normalization method according to an illustrative embodiment of the invention. The spectroscopic system includes a console 102 connected to a probe 104 by way of a cable 106, that is depicted in FIG. 2. The cable 106 carries electrical and optical signals between the console 102 and the probe 104. The probe 104 accommodates a disposable component 108 which used only once, and discarded after such use. In one embodiment, the console 102 and the probe 104 are mechanically connected by an articulating arm 110, which can also support the cable 106. The console 102 contains much of the hardware and the software of the system, and the probe 104 contains the necessary hardware for making suitable spectroscopic observations. The details of the system are further explained in conjunction with FIG. 2.

FIG. 2 shows an exemplary operational block diagram 200 of a spectroscopic system of the type depicted in FIG. 1. The spectroscopic system is similar to single-beam spectrometer devices, but is adapted to include the features of the invention. The console 102 includes a computer 202 which executes runs software that controls the operation of the spectroscopic system 100. The software includes one or more modules recorded on machine-readable media, which can be any medium such as magnetic disks, magnetic tape, CD-ROM, semiconductor memory, or the like. Preferably, the machine-readable medium is resident within the computer

202. In alternative embodiments, the machine-readable medium can be connected to the computer 202 by a communication link. In alternative embodiments, one can substitute computer insructions in the form of hardwired logic for software, or one can substitute firmware (i.e., computer instructions recorded on devices such as PROMs, EPROMS oe EEPROMs, or the like) for software. The term machine –readable instructions as used herein is intended to encompass software, hardwired logic, firmware and the like.

The computer 202 is a general purpose computer. The computer 202 can be an embedded computer, or a personal computer such as a laptop or desktop computer, that is capable of running the software, issuing suitable control commands, and recording information in real time. The computer 202 has a display 204 for reporting information to an operator of the spectroscopic system 100, a keyboard 206 for enabling the operator to enter information and commands, and a printer 208 for providing a print-out, or permanent record, of measurements made by the spectroscopic system 100 and for printing diagnostic results, for example, for inclusion in the chart of a patient. As described below in more detail, in an illustrative embodiment of the invention, some commands entered at the keyboard, enable a user to select particular segments of a spectrum for normalization. Other commands enable a user to select the wavelength range for each particular segment and to specify both wavelength contiguous and non-contiguous segments.

The console 102 also includes an ultraviolet (UV) source 210 such as a nitrogen laser or a frequency-tripled Nd:YAG laser, a white light source 212 such as one or more Xenon flash lamps, and control electronics 214 for controlling the light sources both as to intensity and as to the time of onset of operation and the duration of operation. One or more power supplies 216 are

included in the console 102, to provide regulated power for the operation of all of the components. The console 102 also includes at least one spectrometer and at least one detector (spectrometer and detector 218) suitable for use with each of the light sources. In some embodiments, a single spectrometer can operate with both the UV light source and the white light source. In some embodiments, the same detector can record UV and white light signals, and in some embodiments different detectors are used for each light source.

The console 102 also includes coupling optics 220 to couple the UV illumination from the UV light source 210 to one or more optical fibers in the cable 106 for transmission to the probe 104, and for coupling the white light illumination from the white light source 212 to one or more optical fibers in the cable 106 for transmission to the probe 104. The console 102 also includes coupling optics 222 to couple the spectral response of a specimen to UV illumination from the UV light source 210 observed by the probe 104 and carried by one or more optical fibers in the cable 106 for transmission to the spectrometer and detector 218, and for coupling the spectral response of a specimen to the white light illumination from the white light source 212 observed by the probe 104 and carried by one or more optical fibers in the cable 106 for transmission to the spectrometer and detector 218. The console 102 includes a footswitch 224 to enable an operator of the spectroscopic system 100 to signal when it is appropriate to commence a spectral observation by stepping on the switch. In this manner, the operator has his or her hands free to perform other tasks, for example, aligning the probe 104.

The console 102 includes a calibration port 226 for calibrating the optical components of the spectrometer system. The operator places the probe 104 in registry with the calibration port 226 and issues a command that starts the calibration operation. In the calibration operation, a

calibrated light source provides illumination of known intensity as a function of wavelength as a calibration signal. The probe 104 detects the calibration signal. The probe 104 transmits the detected signal through the optical fiber in the cable 106, through the coupling optics 222 to the spectrometer and detector 218. A test spectral result is obtained. A calibration of the spectral system is computed as the ratio of the amplitude of the known illumination at a particular wavelength divided by the test spectral result at the same wavelength.

The probe 104 includes probe optics 230 for illuminating a specimen to be analyzed with UV and white light from the UV source 210 and the white light source 212, and for collecting the fluorescent and backscatter illumination from the specimen that is being analyzed. The probe includes a scanner assembly 232 that provides illumination from the UV source 210 in a raster pattern over a target area of the specimen of cervical tissue to be analyzed. The probe includes a video camera 234 for observing and recording visual images of the specimen under analysis. The probe 104 includes a targeting souce 236, which can be used to determine where on the surface of the specimen to be analyzed the probe 104 is pointing. The probe 104 also includes a white light illuminator 238 to assist the operator in visualizing the specimen to be analyzed. Once the operator aligns the spectroscopic system and depresses the footswitch 224, the computer 202 controls the actions of the light sources 210, 212, the coupling optics 220, the transmission of light signals and electrical signals through the cable 106, the operation of the probe optics 230 and the scanner assembly 232, the retreival of observed spectra via the cable 106, the coupling of the observed spectra via the coupling optics 222 into the spectrometer and detector 218, the operation of the spectrometer and detector 218, and the subsequent signal procesing and analysis of the recorded spectra.

FIG. 3 is a schematic flow diagram 300 of an illustrative spectral analysis process incorporating features of the invention. In FIG. 3, the flow of information for both the fluorescence spectra and the broadband reflectance spectra is explained in overview. FIG. 3 indicates that the computer 202 has procesed one or more fluorescence spectra to the point where there is a suitable set of spectral results for analysis. With respect to fluorescence data, the illustrative analysis of FIG. 3 includes reading data 302, calibrating the data 304, pre-processing the data 306 and qualifying the data 308 as acceptable, valid data. With respect to the white light broadband reflectance spectra, the illustrative analysis of FIG. 3 includes reading the data 302', calibrating the data 304', pre-processing the data 306', and qualifying the data 308' as acceptable, valid data. The computer 202 combines the data obtained from the fluorescence spectra and the data obtained from the white light broadband reflectance spectra to classify the specimen in a classification step 310. As necessary, the spectroscopic system 100 generates an image from the two types of spectral data, and provides the image as output 312 in a form desired by the colposcopist/user, in either or both of a real-time video image or a recorded image in printed and/or electronic form.

FIG. 4 is a more detailed schematic flow diagram 400 depicting an exemplary calibration step 304, 304' of the type depicted in FIG. 3 and incorporating a non-uniform segment normalization method 430 according to an illustrative embodiment of the invention. In step 410, the illustrative spectroscopic system 100 performs a check of the quality of the spectrum, for example, by examining the signal-to-noise ratio of the spectrum to insure that the spectrum is of acceptable quality. In step 415, in response to the result of the check of step 410 showing an unacceptable spectral quality, the process 400 rejects the spectrum. As indicated in step 420, in

response to the check of step 410 showing a sufficient spectral quality, the process 400 accepts the spectrum. As shown in step 430, the process 400 normalizes acceptable spectra using the non-uniform segment normalization method of the invention. As depicted in step 440, the process 400 records the normalized spectrum for further processing and analysis.

FIG. 5 is a more detailed flow diagram 500 showing an illustrative non-uniform segment normalization method according to the invention. In one embodiment, a detector detects the spectrum as amplitudes as a function of wavelength from the spectrometer. In a further embodiment, an analog-to-digital converter (A/D converter) converts the amplitude of the spectrum at each discrete wavelength to a digital value. The A/D converter provides output at a desired precision, such as 10-bits, 12-bits, 14-bits or even higher precision. As indicated at step 510, the digitized amplitudes so obtained are recorded in a computer memory or machine-readable record as a table of amplitudes recorded at selected discrete wavelengths. The computer 202 divides the spectrum into a plurality of portions, or segments of the spectrum, selected to span a range comprising one or more wavelengths. Each segment is bounded by an upper wavelength and a lower wavelength. As indicated at step 520, the computer 202 selects a subset of the plurality of segments for the normalization process. Alternatively, the user may select the subset of segments for normalization. The segments or ranges do not have to be uniform in width in wavelength space, nor do the segments need to be contiguous with each other or evenly separated in wavelength space. However, the segments may have uniform width, be contiguous and/or be evenly separated in wavelength space.. The width of a wavelength range is defined as the absolute magnitude of a difference between an upper wavelength and a lower wavelength.

Since the ranges and separations can be non-uniform in wavelength space, the method is referred to herein as the "non-uniform segment" normalization method.

The area of each segment is computed. As denoted in step 530, the area is calculated by summing a number n of strips. Each strip has an area determined by multiplying the amplitude at the particuar wavelength corresponding to the strip by a range of wavelengths extending from that wavelength to the next longer discrete wavelength in the spectrum. This integration is expressed mathematically as

$$A_i = \sum_{j=1}^{n} S(\lambda_j)\left|\lambda_{j+1} - \lambda_j\right|$$

where A_i is the area of the i^{th} segment, there are n amplitudes in the i^{th} segment represented by a series of intensities or amplitudes at specific wavelengths, with n corresponding wavelength ranges, the amplitude S at wavelength λ_i being denoted by $S(\lambda_i)$, and the difference $\lambda_{j+1} - \lambda_j$ representing the distance along the wavelength axis between successive amplitudes. This computation is also known as numerical integration. Skilled artisans will appreciate that any one of a number of methods may be used to determine the area of the segments of interest.

In step 540, a total area for all of the segments is determined by summing the values of the A_i. As shown in step 550, the computer 202 sums the areas A_i to obtain the value NF, where NF is the normalization factor, and normalizes each of the amplitudes $S(\lambda_i)$ by dividing by the value NF, or by multiplying by the reciprocal of NF. The computer 202 obtains a normalized set of amplitudes. In step 560, the computer 202 records this normalized set of amplitudes as a new table of amplitude vs. wavelength. The normalized spectrum is denoted as: $\mathbf{S}_A = \frac{1}{NF}\mathbf{S}$. The

computer 202 uses the normalized spectrum to determine a state of health or disease for the tissue specimen being examined.

FIG. 6 is a graph 600 depicting a selection of particular wavelength regions for applying a non-uniform segment normalization process according to the invention. In one embodiment, the specimen is illuminated with ultraviolet radiation of a wavelength of 355 nm from a frequency-tripled Nd:YAG laser.

As depicted in FIG.6, fluorescence spectra are recorded from various tissue specimens having different disease states, or different states of health. The spectrum 602 is typical of healthy cervical tissue comprising normal squamous cells. Cervical tissue can exhibit pre-cancerous lesions known as cervical intraepithelial neoplasia (CIN), a condition that has been divided into three grades. CIN I is the mildest form of the neoplasia, and most often regresses to normal tissue without intervention. CIN II and CIN III are more severe grades of the neoplasia, with CIN III being a potential signal for progression into carcinoma in-situ (CIS). Often the course of treatments for CIN II and CIN III are similar, including removal of the tissue through biopsy or Loop Electro-Excision Procedure (LEEP), so pathologists usually combine the diagnosis of CIN II and CIN III together as CIN II/III. The spectrum 604 in FIG. 6 is typical of CIN II/III. The spectrum 606 is typical of CIN I.

In FIG. 6, two regions are indicated by vertical lines that intersect the three spectra 602, 604, 606. These vertical lines define two regions, one labeled R1, extending from the wavelengths 414 nm to 451 nm inclusive, and another labeled R2, extending from 452 nm to 489 nm inclusive. In one embodiment of the non-uniform segment normalization method, these regions are selected. These spectral regions provide data that most readily distinguish the

conditions of normal health and of CIN II/III. It is not helpful to use spectral information that lies outside these two regions. Such use risks introducing artifacts that may render it more difficult to discriminiate between the conditions of normal health and CIN II/III. For example, in the spectra of FIG. 6, each spectrum 602, 604, 606 exhibits a maximum value at a wavelength of approximately 532 nm, which is outside the range of interest. Peak normalization using the 532 nm line runs the risk that an error in a non-meaningful datum can skew the data in the region of interest. Alternatively, area normalization using the area of each spectrum over the entire spectral range of apprximately 370 nm to approximately 600 nm also runs the risk of normalizing the useful data using an area heavily influenced by non-meaningful data.

According to an illustrative embodiment of the invention, the computer 202 computes a test parameter, for example, the average value of the normalized amplitude within each region that has been selected. In the embodiment described in FIG. 6, the computer 202 can, for example, compute the average normalized amplitude for region R1 and the average normalized amplitude for region R2.

The computer 202 uses the spectra normalized using the non-uniform segment normalization to determine a disease state or a state of health of the specimen from which they were recorded (i.e., the test specimen). In one embodiment, the computer 202 performs the analysis by comparing the spectra obtained from the test specimen to spectra obtained from known healthy and diseased specimens (i.e., known spectra). The computer 202 determines which known spectrum the spectrum obtained from the test specimen most closely resembles. In the embodiment described in FIG. 6, the computer 202 can compare the test parameters

computed by finding the average normalized amplitude within region R1 and within region R2 to the same parameters computed for specimens of known health status.

Equivalents

While the invention has been particularly shown and described with reference to specific preferred embodiments, it should be understood by skilled artisans that various changes in form and detail may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

What is claimed is: